INVESTIGATION OF AN ELASTIN-LIKE POLYPEPTIDE TAG LENGTH

An Honors Thesis Presented in Fulfillment of the Requirements for Graduation with Distinction in Chemical and Biomolecular Engineering at The Ohio State University

By Robert Thomas Wensing

******

The Ohio State University
2011

Oral Examination Committee:
Dr. David Wood
Dr. James Rathman

ABSTRACT

Recombinant protein technology can be used to express a chimeric protein fusion in prokaryotic cells to yield a purified single protein product. By using the elastin-like polypeptide (ELP) affinity tag technology in conjunction with a self-cleaving protein domain from a bacterial intein element, the target protein can be purified using only a change in salt concentration and pH. Shortening the affinity-purification peptide tag has the potential to increase expression by freeing more of the cells energy to produce the valuable protein product. The reduction in tag length has the potential of making this novel purification system more industrially applicable by increasing production of the desired product and eliminating the need for costly chromatography columns. The affinity tag length needed may be affected by the size and solubility of the protein product fused with the tag. For this study, different tag sizes from 250 up to the original 550 amino acid length were cloned into pET vectors with two separate target proteins of 255 and 1,024 amino acids long. The pET vectors were expressed in *BLR E. coli* cells and purified using 0.4 M ammonium sulfate. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to qualitatively determine the success of the purification. Activity and Bradford assays were then used to follow quantitatively the success of the purification. A minimum of a 27 % reduction in tag length was shown to uphold successful purification, and was dependent on the fused protein. Target protein size however was shown not to be the dominating factor in precipitation and success of the purification. More proteins should be looked at to develop a correlation for tag length based on other factors such as hydrophobicity.

Dedicated to the many teachers who have helped me define my academic interests.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. David Wood for his help throughout this project. I would also like to thank my fellow members of the Wood group, with special thanks to Dr. Iraj Ghazi, Dr. Richard Lease, and Michael Coolbaugh.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

Figure Page

1. INTRODUCTION

Over the last two decades the major cost of recombinant protein drug production has shifted from the upstream to the downstream operations. Contained in the upstream costs of recombinant protein drugs is the expression of the desired protein. Downstream costs contain the purification of the protein from other host cell proteins and all other non-protein cell contents. This shift has occurred because of the dramatic increase in achievable fermentation yield because of advancements in expression systems from yielding 20 µg/L to 2 g/L today.[1] Downstream costs have remained high because of the high cost of chromatography resins and affinity tag removal. Table 1 shows the cost, capacity, and regeneration ability of some the most commonly used industrial chromatography resins.[2]

Affinity Technology	Resin	Resin Cost ($/g)	Capacity (µg/mL)	Regeneration Cycles	Manufacturer
Chitin-binding domain	Chitin	180	2	5	NEB
His	Ni-NTA Sepharose 6 Fast Flow	27	40	5	GE
Antibody Fc domain	MabSelect Xtra	3	42	100	GE

Table 1: Industrial affinity resins and associated costs[2]

Adding to the downstream processing cost is the need for the affinity tag to be removed after purification. If the affinity tag is left attached to the protein drug the product may be immunogenic. Affinity tags can be removed through methods such as

protease application or use of inteins.[2] A protease is an enzyme which cleaves at a given amino acid sequence, while an intein is a self-splicing protein which cleaves itself from the amino acids it is attached to. Proteases do not provide a viable solution on an industrial scale because of the large quantities that would be necessary and their inherent biological activity. Because of this biological activity, after a protease cleaves the desired product from the affinity tag an additional purification step must be included to then remove the protease from the product solution.

Inteins provide the most economically viable solution for tag removable because of their ability to be expressed between the affinity tag and protein product. As shown in Figure 1 below, an intein can splice itself out and ligate the extein fragments, or cleave at one end to remove the C terminus extein section while remaining connected to the N-terminus extein.

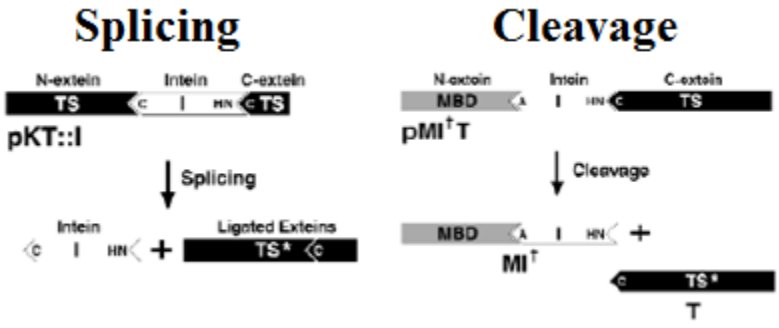

Figure 1: Splicing and cleavage mechanism illustrations of intein variants

The cleaving intein was made possible by mutating the native splicing intein element and provides a valuable intein mutant that allows the protein product to be cleaved from the affinity tag and intein.[3] After binding to a column and addition of a thiol or shift in pH,

the intein is able to selectively release the protein product from the chromatography column.[2] The intein element provides an important capability, to economically cleave the affinity tag, but still does not get around the high cost of chromatography resins.

To avoid using any type of chromatography column and reduce the cost of purification, an elastin-like polypeptide (ELP) affinity tag can be used in conjunction with a self cleaving intein element to purify a recombinant protein using only a shift in salt and pH.[4] An ELP is a repetition of five amino acids V-P-G-X-G, where X is any amino acid. The overall sequence is 550 amino acids in length and hydrophobic. As shown in Figure 2, when temperature or salt concentration of a solution is increased the tag and anything that is fused to it will precipitate out of solution. Originally a temperature of 30° C and 1.5 M NaCl were used to precipitate the ELP tag-intein-target protein fusion out of solution. More recent work has shown that a salt producing stronger kosmotropic ions, shown in Figure 3 on the following page, of ammonium sulfate is able to precipitate an ELP tag-intein-target protein at a much lower concentration of 0.4 M and room temperature.[5]

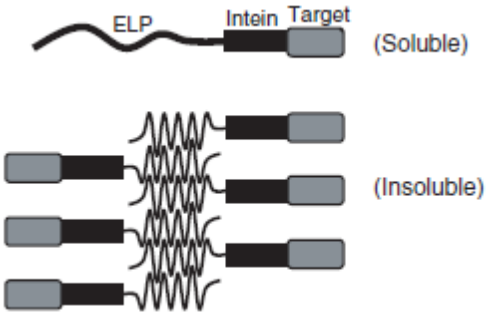

Figure 2: Aggregation of ELP tag at increased temperature or salt concentration

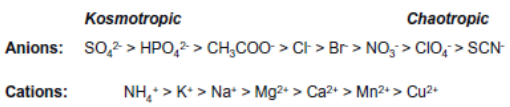

Figure 3: Hofmeister seris of anions and cations[5]

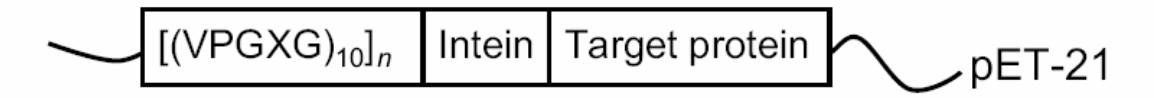

Figure 4: DNA sequence of fused ELP tag, intein, and target protein on a pET plasmid

A target protein can then be purified by expressing the ELP intein sequence in fusion with the target protein.[4] Once the fusion, shown in Figure 4 above, has been over expressed in *Escherichia coli* cells, the cells are lysed and undergo a cycle of centrifugation at 10° C or lower. Centrifugation pellets the cell debris and allows the clarified lysate to be removed as the supernatant. The salt concentration of the clarified lysate is increased and it is brought to room temperature to precipitate the ELP-intein-target protein fusion. After precipitating the fusion protein, the contents are then centrifuged to pellet the precipitate and remove the contaminants as the supernatant. The precipitated fusion protein is then resuspended at a pH of 6.2 to cleave the target protein from the ELP tag and intein. After cleavage, the salt concentration is increased to precipitate the tag and intein while leaving the target protein in solution. The contents are centrifuged once again to pellet the precipitate and recover the purified target protein as the supernatant. This scheme is illustrated graphically in Figure 5 on the following page and is known as the ELP scheme.

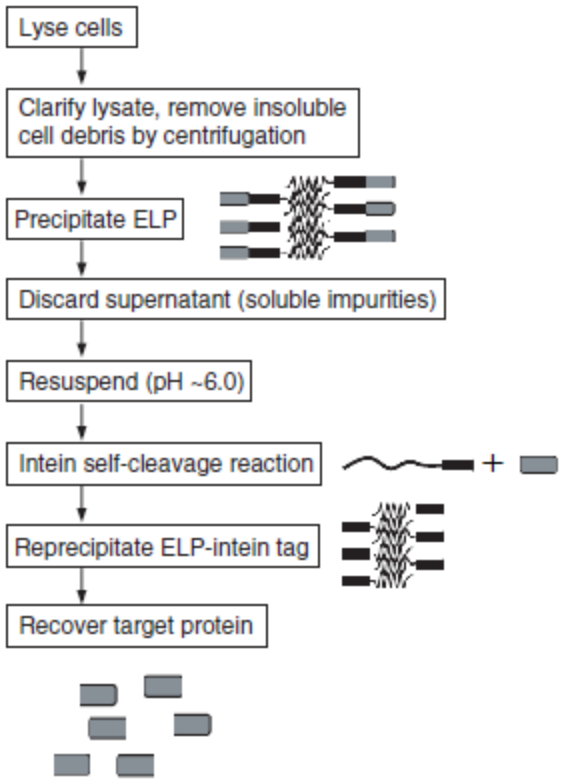

Figure 5: Graphical representation of ELP scheme for protein purification

The ELP scheme has the advantage of removing high cost resins and proteases from downstream processing, but still has the drawback of requiring the expression of a large aggregation tag and intein along with the protein product. In order for this process to become more industrially applicable, a smaller ELP tag length may be able to be introduced in the process to produce an increase in yield. It is hypothesized that a smaller ELP tag will free energy in the cell by relieving the cell from synthesizing the large ELP tag's repetitive protein sequence. By freeing energy in the cell a larger yield of the final protein product may be obtained. Along with looking at reducing the ELP tag length, it will be important for future industrial applications to determine the relationship between the target protein fused with the ELP tag- intein and the tag length required for successful

purification. It is hypothesized that a larger tag length will be needed to successfully purify a larger target protein.

2. MATERIALS AND METHODS

2.1. Construction of Vectors

ELP tag lengths were changed in the DNA sequence using a previously constructed pET vector containing the target protein carbonic anhydrase as shown below in Figure 6. The ELP was reduced from the original 1650 base pairs to 750 base

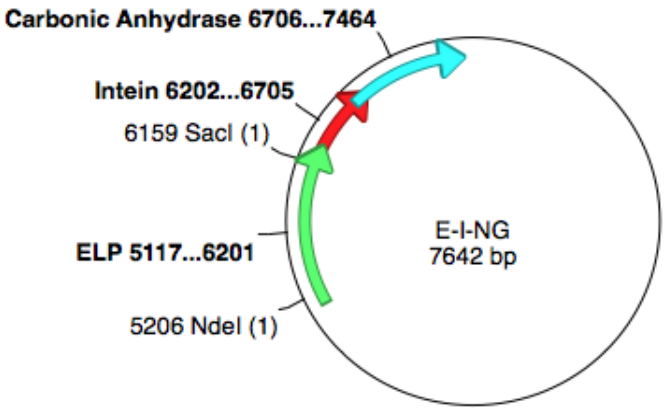

Figure 6: Original pET vector containing 1650 base pair ELP and carbonic anhydrase target protein

pairs in decreasing units of 150 base pairs. The different length ELP sequences of 750, 900, 1050, 1200, 1350, and 1500 base pairs were stored in pUC vectors between the NdeI and SacI restriction enzyme sites as shown in Figure 7 on the following page. Each sequence was isolated by double digesting the pUC vectors with NdeI and SacI

restriction enzymes (New England BioLabs ©). The digested fragments were then run on an agarose gel and gel purified using QIAquick Gel

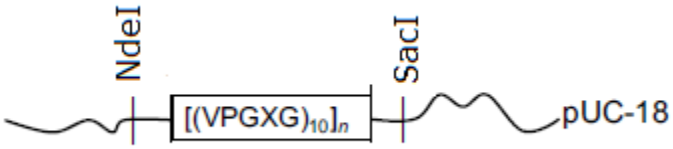

Figure 7: Storage vector of different ELP length sequences

Extraction Kit (Qiagen©). The original pET vector was also digested with the same restriction enzymes, and the vector backbone was isolated using the same purification technique. The ELP inserts were then ligated with 100 ng of backbone in a six-fold molar excess. The ligation was conducted at room temperature for 1 hr. The ligation product was transformed according to the Z-Competent *E. coli* Trasformation Kit protocol (Zymo Research©) using DH5α *E. coli* cells and plated on Luria-Bertani (LB) agar growth media supplemented with ampicillin at a concentration of 20 µg/100 mL LB. The new clones were verified by a digest map check using NdeI and EcoRI restriction enzymes (New England BioLabs ©).

Once vectors containing variable ELP lengths were created, the larger target protein β-galactosidase was cloned into the vectors in place of carbonic anhydrase between the BsrGI and HindIII restriction sites. The β-galactosidase sequence was moved from a similar construct shown in Figure 8 on the next page. The cloning was done as previously described with the exception of using BsrGI and HindIII restriction enzymes

(New England BioLabs ©). The new clones were again verified by a digest map check using NdeI and SacI restriction enzymes.

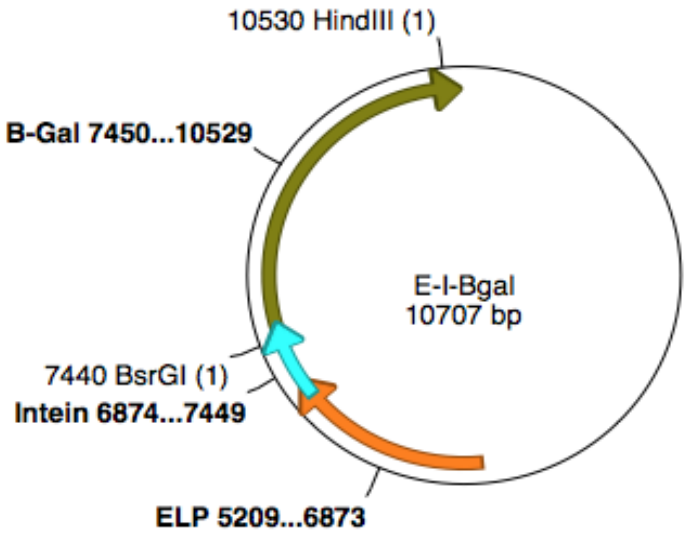

Figure 8: Plasmid where β-galactosidase sequence was originally contained

2.2 Expression of Vectors

Once the correct vectors had been made with variable ELP lengths and different target proteins, the vectors were moved from DH5α *E. coli* cells to BLR *E. coli* cells. All plasmids were transformed in z-competent BLR *E. coli* cells according to the Z-Competent *E. coli* Transformation Kit protocol. The cells were plated on LB agar media supplemented with ampicillin at a concentration of 20 µg/100 mL LB. Plates were incubated at 37° C for 16 hours. A single colony from each plate was selected to inoculate 5 mL of LB media supplemented with 0.1 µg of ampicillin. The inoculated cultures were continuously shaken and incubated overnight at 37° C for 16 hours. One mL of overnight culture was then used to inoculate 100 mL of Terrific Broth (TB)

supplemented with 20 µg of ampicillin. The culture was then placed to shake continuously at 37° C for 4 hours, and then induced with 600 µL of 100mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). For cultures containing carbonic anhydrase, 300 µL of 100 mM zinc sulfate was added. The cultures were then continuously shaken and placed at 20° C for an additional 20 hours.

2.3 Cell Harvesting and Protein Purification

After over expression in BLR *E. coli* cells, the cells were harvested using centrifugation and the target proteins were purified following the ELP scheme. Each culture was centrifuged for 10 minutes at 4,000g and 4° C directly after expression. The supernatant was immediately removed and cell pellets were resuspended with 10 mL/g pellet in lysis buffer (10 mM Tris-HCl pH 8.65, 2 mM EDTA). EDTA was not included in the lysis buffer for expressions containing carbonic anhydrase. After the pellets were resuspended in lysis buffer, the mixtures were divided into 1 mL aliquots in eppendorf tubes and frozen at -20° C.

For all purifications, 1 mL samples were taken from -20° C and thawed. Each thawed sample was sonicated for 6 pulses of 15 s on a setting of 4-5 W. A 100 µL aliquot of the whole lysate was taken from each tube for SDS-PAGE analysis, activity assay, and Bradford assay. The samples were then centrifuged at 15,000 rpm for 10 min on a tabletop centrifuge and the supernatant was recovered. A 90 µL aliquot of clarified lysate was taken from each tube for SDS-PAGE analysis, activity assay, and Bradford assay. Six hundred microliters of clarified lysate were mixed with 600 µL of 0.8 M $(NH_4)_2SO_4$ and incubated at room temperature for 30 min. The samples were then centrifuged at

15,000 rpm for 10 min on a table top centrifuge and the supernatant was removed as the soluble contaminants. The soluble contaminants were kept for SDS-PAGE analysis, activity assay, and Bradford assay. The pellets were then washed with 250 mL of 0.4 M $(NH_4)_2SO_4$ and centrifuged at 15,000 rpm for 10 min on a tabletop centrifuge after which the supernatant was removed and discarded. The pellets were then resuspended in 400 µL of cleaving buffer (40 mM Tris-HCl pH 6.2) and left at room temperature for 20 hr. Samples of 40 µL were taken immediately after addition of cleaving buffer and after the 20 hr cleavage period. Samples were incubated at room temperature for 30 min after addition of 320 µL of 0.8 M $(NH_4)_2SO_4$. The samples were then centrifuged at 15,000 rpm for 10 min on a tabletop centrifuge. The supernatant was recovered as the product for SDS-PAGE analysis, activity assay, and Bradford assay.

2.4. SDS-PAGE, Activity, and Bradford Analysis

For each SDS-PAGE sample, 40 µL of experiment sample was mixed with 40 µL sample buffer (950 µL Laemmli Sample Buffer (Bio-Rad©), 50 µL β-mercaptoethanol) and heated at 95 ° C for 5 min. The SDS-PAGE samples were then run on a 5% polyacrylamide gel at 25 mA.

The enzyme activity of carbonic anhydrase was measured using the colorimetric shift from red to yellow of phenol red indicator when a pH shift from 8.3 to 6.3 occurs in a 20 mM tris-sulfate buffered solution.[6] Carbon dioxide is catalyzed in the presence of carbonic anhydrase to carbonic acid effectively dropping the pH. The assay was performed at 0° C and the activity was computed in Wilbur-Anderson (W-A) units. One

W-A unit corresponds to dropping the pH of a 0.02 M Trizma buffer from 8.3 to 6.3 per minute at 0° C.

The activity of β-galactosidase was measured using a colorimetric change that corresponded to the hydrolysis of o-nitrophenyl-β-D-galactopyranoside (ONPG).[7] The enzyme β-galactosidase catalyzes the hydrolysis of galactosides such as ONPG. The substrate ONPG is converted to galactose and the chromophore *o*-nitrophenol. The chromophore *o*-nitrophenol has a peack absorbance at a wave length of 420 nm. The assay was performed in accordance with the β-gal Activity Assay Kit (Stratagene©). The absorbance was read on a 96 well microplate reader.

Protein concentration was measured using a Bradford Assay by measuringabsorbance at 595 nm. Protein standards of 0.2, 0.4, 0.46, 0.8, 0.9 µg Bovine serum albumin(BSA)/mL H_2O were used for calibration. An appropriately diluted sample of 20 µL was added to 980 µL of Coomassie Brilliant Blue G-250 dye (Bio-Rad©) and incubated for 5 min. The absorbance was then read at 595 nm and concentration was determined according to the previous calibration.

3. RESULTS

SDS-PAGE analysis was performed at each stage of purification for the 255 amino acid carbonic anhydrase samples. When using the original 550 amino acid length ELP tag in the ELP scheme, as shown on the next page in Figure 9, the majority of proteins were removed in the soluble contaminants and the purified product was recovered in the final supernatant. A slight contaminant is apparent just above the target protein product shown in product lane f of Figure 9.

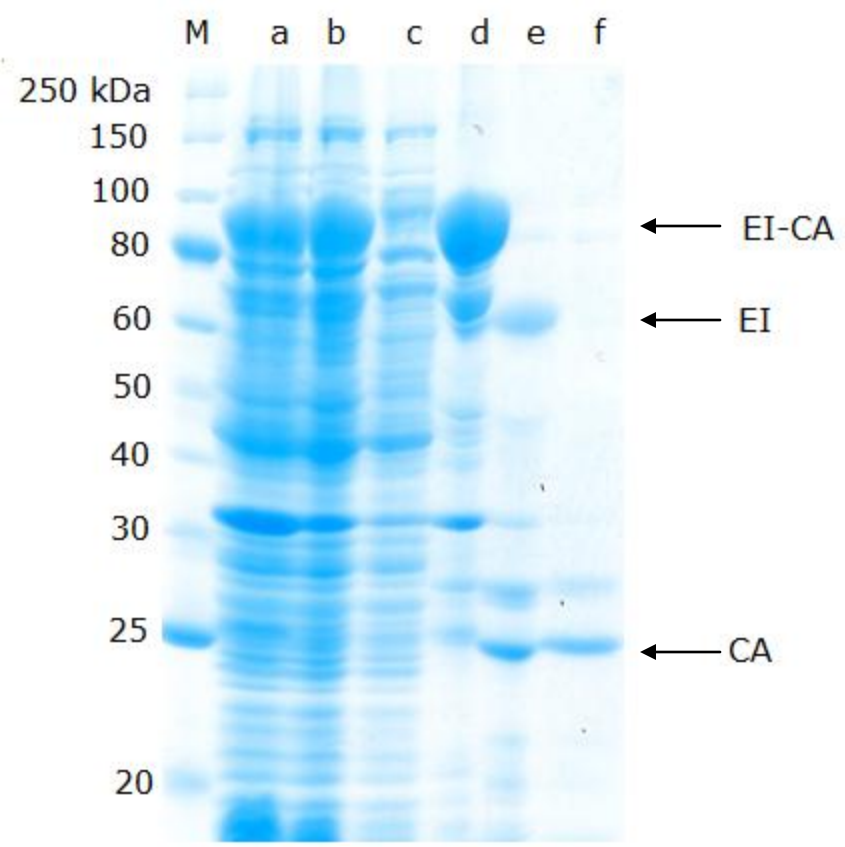

Figure 9: SDS-PAGE of all purification steps for the original ELP length of 550 amino acids with carbonic anhydrase as the target protein. Lanes: M: molecular weight marker with the known protein sizes indicated on the left side of the gel, a: whole lysate, b: clarified lysate, c: soluble contaminants, d: start of the cleavage reaction where the E-CA is still connected, e: end of the cleavage reaction where the ELP and intein (EI) and carbonic anhydrase (CA) have split, f: purified product.

Figure 10 on the following page shows an SDS-PAGE gel of the carbonic anhydrase products after ELP scheme purification with shortened ELP tag lengths. No target protein is seen when using an ELP tag length of 300 amino acids, and the target

protein does not reach full density until an ELP tag length of 400 amino acids is used. In lanes a and b in Figure 10, the EI cleaved tag and intein are visible in the product between 40 and 50 kDa. The EI tag is slightly larger in lane b because of the additional 50 amino acids the ELP tag contains.

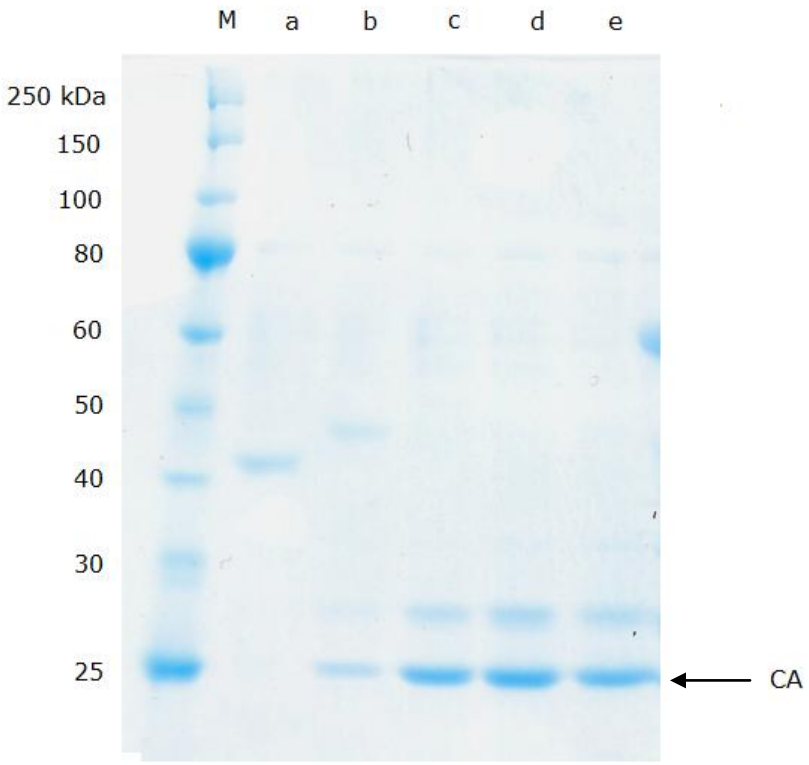

Figure 10: SDS-PAGE analysis of purification products using various ELP tag lengths and target protein carbonic anhydrase. Lanes: M: molecular weight marker with the known protein sizes indicated on the left side of the gel, a: ELP tag 300 amino acids, b: ELP tag 350 amino acids, c: ELP tag 400 amino acids, d: ELP 450 amino acids, e: ELP 550 amino acids.

As shown in Table 2 below, the activity of the soluble contaminants increases at an ELP tag size lower than 400 amino acids and the activity of the product increases significantly over 350 amino acids. The purification factor when using a 350 amino acid tag is approximately half that of using a 400 amino acid tag. This result corresponds well with the gel result seen in Figure 10 previously. The yield of protein shows a decrease when using an ELP tag less than 400 amino acids. It is important to note that the 300 and 350 amino acid ELP purifications did not result in a completely purified protein and yield data is based off total protein in the sample.

ELP Tag Length (amino acids)	Quantity of Purified Protein (µg)	Activity of Soluble Contaminants (U/ml)	Activity of Product (U/mL)	Purification Factor
300	0.004	5940	3340	1.50
350	0.006	6160	19630	9.09
400	0.010	3520	44640	20.36
450	0.016	2500	28360	14.76
550	0.012	2270	28100	16.05

Table 2: Quantitative analysis of carbonic anhydrase purifications with various ELP tag lengths

Figure 11 on the next page shows an SDS-PAGE gel of the β-galactosidase products after ELP scheme purification with shortened ELP tag lengths. All purifications using shortened ELP tag lengths resulted in recovered target protein. In lanes a, b, and c in Figure 11, the EI cleaved tag and intein are visible in the product between 40 and 50 kDa. The EI tag is slightly larger in lanes b and c because of the additional 50 and 100 amino acids, respectively, the ELP tag contains.

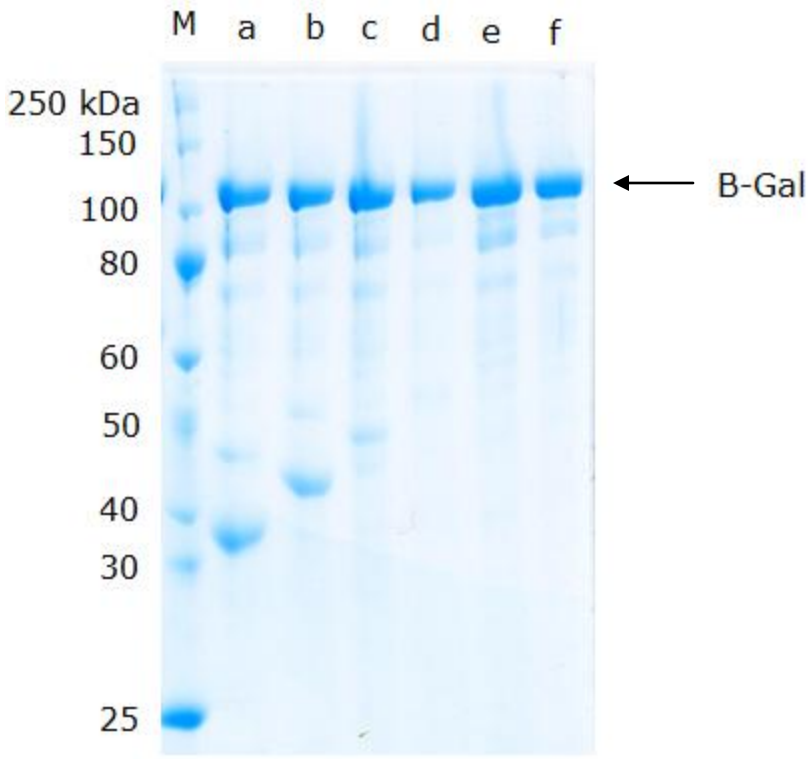

Figure 11: SDS-PAGE analysis of purification products using various ELP tag lengths and target protein β-galactosidase. Lanes: M: molecular weight marker with the known protein sizes indicated on the left side of the gel, a:ELP tag 250 amino acids b: ELP tag 300 amino acids, c: ELP tag 350 amino acids, d: ELP tag 400 amino acids, e: ELP 450 amino acids, f: ELP 550 amino acids.

As shown in Table 3 on the next page the activity of the soluble contaminants increases with decreasing tag length, and the product activity remains constant regardless of the ELP tag length used. The purification factor also remains approximately constant, corresponding well with the gel result in Figure 11. The total yield of protein was slightly higher for the lower ELP tag lengths of 250, 300, and 350 amino acids. Again it is

important to note these purifications did not result in a completely purified product and yield is based off total protein in the sample.

ELP Tag Length (amino acids)	Quantity of Purified Protein (μg)	Activity of Soluble Contaminants (U/ml)	Activity of Product (U/mL)	Purification Factor
250	0.030	61565.22	125993.49	2.40
300	0.030	27053.14	102523.07	1.75
350	0.033	16347.83	165146.42	2.98
400	0.021	11130.43	90681.18	3.16
450	0.026	6067.63	121939.51	3.06
550	0.022	4173.91	84706.89	2.23

Table 3: Quantitative analysis using Bradford and activity assay results

Based on the SDS-PAGE results and further supported by activity and Bradford assays, carbonic anhydrase with a length of 255 amino acids was successfully purified with a minimum ELP tag length of 400 amino acids, and β-galactosidase with a length of 1,024 amino acids was successfully purified with a minimum ELP tag length of 250 amino acids.

4. DISCUSSION AND CONCLUSIONS

The ELP tag size required was reduced from the original 550 amino acids to 400 amino acids. This reduction in tag size did not correspond with any significant increase in product yield. However, even without the existence of an increase in product yield the smaller tag size would still free the cell's energy from making a non-valuable ELP tag. Although an increase in product yield was not seen at this experimental scale there may be an increase in yield if this process is scaled to an industrial production level.

The ratio of ELP tag length needed to target protein size contradicted the original hypothesis of a longer ELP tag length being necessary for a larger protein. Although less than 20% the weight of β-galactosidase, carbonic anhydrase required a 60% larger length ELP tag to be successfully purified. Thus, target protein size is not the dominating factor in determining the ELP tag length needed for successful purification. Other factors such as quaternary structure and hydrophobicity could be investigated in the future as potential dominating factors. β-Galactosidase is a tetramer; the quaternary structure may be forcing ELP tags on nearby subunits to interact. The hydrophobicity of the target protein may also be a major determining factor in the precipitation since the ELP tag itself precipitates because it is hydrophobic.

Although ELP tag lengths of less than 400 amino acids led to a successful recovery of β-galactosidase the product was not pure. At the lower ELP tag lengths the EI cleaved ELP tag and intein showed up in the product lane. Once cleaved, the precipitation and removal of the cleaved ELP and intein is not determined by the target protein. For tag sizes less than 400 amino acids to be used a higher $(NH_4)_2SO_4$ concentration is necessary to fully precipitate the cleaved ELP tag and intein.

The ELP tag length needed for successful purification has been reduced by 27% and has the potential of being decreased even more depending on the target protein. Future work should be directed at defining a larger design space for ELP tag purifications. This larger design space should include the use of increased salt concentrations, different levels of hydrophobic and hydrophilic proteins, and different conformations and levels of quaternary structure.

REFERENCES

[1]Werner, R. G. Economic aspects of commercial manufacture of biopharmaceuticals, *J Biotechnol* **113** (2004), pp. 171–182.

[2]Fong B.A., Wu, W., Wood, D.W. The potential role of self-cleaving purification tags in commercial-scale processes. *Protein Expr. Purif.* **66,** 198–202(2009).

[3]Wood, D.W., Wu, W., Belfort, G., Derbyshire, V., Belfort, M. A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* **17**(1999) pp. 889–892

[4] Banki, M. R., Feng, L., Wood, D. W. Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nature Online Publications.* (2005) pp. 1-3

[5]Fong, B.A., Wu, W.Y. *and* Wood, D.W. Optimization of ELP-intein mediated protein purification by salt substitution. *Protein Expr. Purif.* **66** *(*2009) pp. 198–202.

[6]Sigma-Aldrich. Enzymatic Assay of Carbonic Anhydrase for Wilbur-Anderson Units. http://www.sigmaaldrich.com/sigma-aldrich/technical-documents/protocols/biology/enzymatic-assay-of43.html (accessed July 20, 2010).

[7]Craven, G.R., Steers, E., Jr. & Anfinsen, C.B. Purification, Composition, and Molecular Weight of the Beta-galactosidase of Escherichia Coli K12. *J Biol Chem* **240** (1965) pp. 2468-2477